

Efficient wavefunction propagation by minimizing accumulated action

Zachary B. Walters¹

¹*Max Born Institute, Bereich B, Max Born Strasse 2a, 12489 Berlin Adlershof, Berlin Germany*
(Dated: August 9, 2011)

This paper presents a new technique to calculate the evolution of a quantum wavefunction in a chosen spatial basis by minimizing the accumulated action. Introduction of a finite temporal basis reduces the problem to a set of linear equations, while an appropriate choice of temporal basis set offers improved convergence relative to methods based on matrix exponentiation for a class of physically relevant problems.

Calculating the time evolution of a quantum wavefunction is a longstanding problem in computational physics. The fundamental nature of the problem makes it resistant to simplification. Problems of physical interest, such as the interaction of a molecule with a strong laser field, may lack symmetry or involve time dependent, nonperturbative fields. Problems involving multiple dimensions or multiple interacting particles may quickly grow so large as to be unmanageable with all but the largest computational resources [1–3], a problem which is not easily outstripped by increases in computational power. An ideal propagator, then, must serve two masters – it must treat the physical side of the problem accurately, and the computational side of the problem efficiently.

The most common approach to the problem does not treat the evolving wavefunction directly. Instead, the wavefunction $\psi(x, t)$ is expanded in some chosen basis set $\phi(x, t) = \sum_i c_i(t) \chi_i(x)$, where $c_i(t_0) = \langle \chi_i | \psi(t_0) \rangle$. Here $\psi(x, t)$ is the true wavefunction and $\phi(x, t)$ is its representation in the chosen basis. After this expansion has been made, the propagation scheme may operate only on the wavefunction's representation, rather than the wavefunction itself. Within this general framework, there has been a great profusion of methods for calculating the evolution of the coefficients $c_i(t)$. Popular methods include Crank-Nicholson[4], second order differencing[5], split operator[6], short iterative Lanczos[7], and Chebyshev propagation[8], as well as many others[9].

As may be inferred from the large number of competing methods, the practical question of which method works best is very difficult to answer, and usually requires problem-specific information. Ironically, the time dependent Schrödinger equation (TDSE), while challenging to solve, is not difficult to satisfy at a particular time. All of the above methods satisfy the TDSE exactly or approximately at the initial time in the propagation interval. Indeed, entire families of propagators may satisfy the TDSE at the initial point: for the class of propagators $\psi(x, t)(1 - iH\alpha\Delta t) = \psi(x, t + \Delta t)(1 + iH(1 - \alpha)\Delta t)$, $0 \leq \Delta t$, the TDSE at time t is satisfied to first order for any value of α . $\alpha = 1$ yields the forwards Euler method, $\alpha = 0$ the backwards Euler, $\alpha = 1/2$ the Crank-Nicholson. Although such methods may show sharp differences in suitability to particular problems, the TDSE

alone gives little guidance. A full diagonalization of the Hamiltonian matrix would satisfy the TDSE at all times; however, such a diagonalization would be prohibitively expensive for a large problem, and would not exist for a problem with a time dependent Hamiltonian. Comparison of different propagators has often involved numerical testing on simple problems[10, 11] or algorithmic scaling arguments[12].

This paper addresses the problem of wavefunction propagation from the physical perspective of minimizing the action accumulated over the chosen time interval. Minimizing this action is shown to be equivalent to minimizing the time integrated error of propagation. Because the action is calculated over the entire time step rather than at a single point, the constraint that it be minimized is more strict than the TDSE, allowing the construction of a unique, variationally optimum propagator for a particular order in time.

ERRORS OF PROPAGATION AND REPRESENTATION

A central difficulty of any numerical propagation scheme is that, although the propagator seeks to model the evolution of an ideal wavefunction $\psi(x, t)$, it has access only to the representation of the wavefunction in some chosen basis, $\phi(x, t)$. The error of the representation is given by $\delta(x, t) = \psi(x, t) - \phi(x, t)$.

As an alternative to direct exponentiation of the Hamiltonian, a propagator may be constructed by minimizing the integral of the error over the time step

$$\text{Global error} = \int_t^{t+\Delta t} dt' \langle \delta(x, t') | \delta(x, t') \rangle. \quad (1)$$

Writing the error as a two term Taylor series,

$$\begin{aligned} \text{Global error} = & \int_t^{t+\Delta t} dt' \langle \delta(x, t) | \delta(x, t) \rangle + \\ & \int_t^{t+\Delta t} dt'' \int_t^{t'} \frac{d}{dt'} \langle \delta(x, t') | \delta(x, t') \rangle, \end{aligned} \quad (2)$$

and recalling that $i \frac{d}{dt} \psi(x, t) = H \psi(x, t)$ for the true wavefunction, the second term in equation 2 can be writ-

ten as

$$\begin{aligned} \frac{d}{dt} \langle \delta(x, t) | \delta(x, t) \rangle &= 2i \langle \phi(x, t) | (i \frac{d}{dt} - H) | \phi(x, t) \rangle + \\ &2i \langle \delta(x, t) | H | \delta(x, t) \rangle. \end{aligned} \quad (3)$$

From equations 2 and 3, it is apparent that the global error may arise either from imperfectly representing the wavefunction in a particular basis (terms containing $\delta(x, t)$) or from imperfectly describing the evolution of the wavefunction in that basis (terms containing $\phi(x, t)$). The representation error may be minimized by an appropriate choice of basis; here the focus is on minimizing the error of propagation.

The quantity $\langle \phi(x, t) | (i \frac{d}{dt} - H) | \phi(x, t) \rangle$ found in equation 3 is the Lagrangian density, and its integral over time gives the action accumulated by the wavefunction in a particular interval. However, unlike the true Lagrangian density, here the action is calculated with respect to the representation of the wavefunction, rather than the wavefunction itself. The distinction is significant. For the true wavefunction, minimizing the action is equivalent to setting $i \frac{d}{dt} \psi(x, t) - H \psi(x, t) = 0$ for all x and t . For the action minimizing representation of the wavefunction, $|\langle \phi(x, t) | (i \frac{d}{dt} - H) | \phi(x, t) \rangle|$ is dependent on the choice of spatial and temporal basis functions and is not guaranteed to be zero.

For a finite basis of spatial $\chi_i(x)$ and temporal $T_n(t)$ basis functions, a time dependent representation of the wavefunction can be written as

$$\phi(x, t) = \sum_{i,n} C_{in} \chi_i(x) T_n(t). \quad (4)$$

In this basis, the global error of Eq. 3 becomes a function of the coefficients C_{in} and the matrix representations of the quantum mechanical operators. Writing the Hamiltonian as the sum of time independent and time dependent operators

$$H = H_0(x) + V(x, t) \quad (5)$$

and defining the matrices

$$H_{ij} = \int dx \chi_i^*(x) H_0 \chi_j(x) \quad (6)$$

$$V_{ijnm} = \int_t^{t+\Delta t} dt' \int dx \chi_i^*(x) T_n^*(t') V(x, t') \chi_j(x) T_m(t) \quad (7)$$

$$O_{ij} = \int dx \chi_i^*(x) \chi_j(x) \quad (8)$$

$$U_{nm} = \int dt T_n^*(t) T_m(t) \quad (9)$$

$$Q_{nm} = \int dt T_n^{*'}(t) T_m(t), \quad (10)$$

the change in action resulting from $C_{in} \rightarrow C_{in} + \epsilon_{in}$ is given by

$$\delta S = \sum_{i,j,n,m} C_{in} [i O_{ij} Q_{nm} - H_{ij} U_{nm} - V_{ijnm}] \epsilon_{jm}^* \quad (11)$$

and the condition to minimize the accumulated action is that either $\epsilon_{jm}^* = 0$ (for the initial conditions) or

$$\delta S_{jm} = \sum_{i,n} C_{in} [i O_{ij} Q_{nm} - H_{ij} U_{nm} - V_{ijnm}] \epsilon_{jm}^* = 0 \quad (12)$$

for all j, m . In these equations, i appearing as a subscript is treated as an index, while i multiplying $O_{ij} Q_{nm}$ is the square root of negative one.

Equation 12 is the main result of this paper. In order to construct a least action propagator, it is necessary only to choose an appropriate temporal basis. While in principle this analysis applies equally well to any choice of basis, an obvious choice is for $T_n(t)$ to be a set of linearly independent low-order polynomials in t .

Lagrange interpolating polynomials provide a convenient set of temporal basis functions. For an evenly spaced grid $t_m = t + \Delta t * m/n$ for $m = 0, n$, the interpolating polynomials are given by

$$T_m(t) = \prod_{k=0, n; k \neq m} \frac{t - t_k}{t_m - t_k}. \quad (13)$$

This yields a basis set of n linearly independent n -order polynomials in t , with the property that $\phi(x, t_n) = \sum_i C_{in} \chi_i(x)$. One advantage of this choice of basis is that for small propagation times, C_{in} will have comparable amplitudes for all n , making the associated linear system easier to solve with high accuracy.

Having chosen a temporal basis, coefficients C_{in} which satisfy Equation 12 as well as the initial condition $C_{i0} = \langle \chi_i | \psi(t_0) \rangle$ can be found using Lagrange multipliers. If S is the action accumulated in the time interval, let $S' = S + \sum_i \lambda_i f_i^*$, where $f_i = C_{i0} - \langle \chi_i | \psi(t_0) \rangle$. The least action coefficients are found by minimizing S' with the constraint that $f_i = 0$ for all i . This yields a system of linear equations

$$\sum_{i,n} C_{in} [i O_{ij} Q_{nm} - H_{ij} U_{nm} - V_{ijnm}] + \lambda_j = 0 \quad (14)$$

for all j, m , and

$$C_{i0} = \langle \chi_i | \psi(t_0) \rangle \quad (15)$$

for all i . The Lagrange multipliers λ_j calculated in this procedure are not needed by the propagator and can be discarded after solving the linear system.

COMPARISON WITH THE LANCZOS PROPAGATOR

The least action propagator derived in the previous section is the unique, variationally optimum propagator for a particular order in time. As such, it represents a formal improvement over all propagators approximating the wavefunction as a low order polynomial in time – forward and backwards Euler, Crank Nicholson, second order differencing, etc. However, it is less clear how this formal improvement translates to a practical benefit, or how the least action propagator compares to methods which attempt to diagonalize the Hamiltonian in a Krylov subspace, such as the popular short iterative Lanczos method [7].

The Lanczos method works by repeatedly multiplying the initial wavefunction by the Hamiltonian matrix to create a Krylov space of limited dimension in which the matrix exponential e^{-iHt} can be calculated exactly. It is considered to be both efficient and quickly converging[10]. Existing variational propagators have focused on the evolution of the wavefunction in the Krylov subspace, yielding convergence properties similar to the Lanczos propagator[13, 14]. The Chebyshev propagator, which also uses repeated multiplication by the Hamiltonian matrix to construct a Krylov space, converges similarly to the Lanczos method[15].

In the limit that the Krylov space has dimension equal to the full Hamiltonian, the Lanczos method is equivalent to diagonalizing the Hamiltonian, yielding $L = i\frac{d}{dt}\psi - H\psi = 0$ at all times. This solution is the global minimum action solution, and cannot be improved upon. However, for most applications, the Krylov subspace is chosen to have a much smaller order – typically in the range 1-10.

Because the Krylov subspace is constructed through repeatedly multiplying an initial wavefunction by the Hamiltonian matrix, later Krylov basis vectors will tend to some overlap with those eigenvectors of H with the largest eigenvalues. For problems with Coulomb singularities or fine spatial bases, it is not uncommon for these largest eigenvalues to be artifacts of the choice of basis, having no counterpart in the system being described. However, they may nonetheless serve to limit the stepsize which may be taken with high accuracy.

If an ideal wavefunction (ie, without reference to a basis) $\psi(x, t)$ can be expanded in terms of energy eigenfunctions over a short time interval

$$\psi(x, t) = \sum_{\alpha} c_{\alpha} f_{\alpha}(x) e^{-iE_{\alpha}(t-t_0)} \quad (16)$$

and the evolution of the wavefunction's representation in the Krylov subspace is given by

$$\phi(x, t) = \sum_{\beta} d_{\alpha, \beta} g_{\beta}(x) e^{-iE_{\beta}(t-t_0)}, \quad (17)$$

where $d_{\alpha, \beta} = c_{\alpha} \langle g_{\beta} | f_{\alpha} \rangle$, then the error is given by

$$\delta(x, t) = \sum_{\alpha, \beta} d_{\alpha, \beta} g_{\beta}(x) (e^{-iE_{\alpha}(t-t_0)} - e^{-iE_{\beta}(t-t_0)}) \quad (18)$$

and

$$\langle \delta | \delta \rangle = \sum_{\alpha, \beta} |d_{\alpha, \beta}|^2 4 \sin^2\left(\frac{E_{\alpha} - E_{\beta}}{2}(t - t_0)\right). \quad (19)$$

If the Krylov subspace is now partitioned into a “good” subspace with $E_{\beta} < E_{\text{cutoff}}$ and a “bad” subspace with $E_{\beta} > E_{\text{cutoff}}$, the error can be estimated by setting $E_{\alpha} - E_{\beta} = 0$ in the good subspace and $E_{\beta} - E_{\alpha} = E_H$, where E_H reflects the largest eigenvalues of H , yielding

$$\langle \delta | \delta \rangle \approx \sum_{\alpha, \beta, E_{\beta} > E_{\text{cutoff}}} |d_{\alpha, \beta}|^2 4 \sin^2\left(\frac{1}{2}E_H(t - t_0)\right). \quad (20)$$

The error of the Lanczos method is thus minimized either when the projection into the bad subspace is small or when $E_H \Delta t \ll 1$.

In contrast to the Lanczos method, the error of the least action propagator is bounded by the error of the Taylor series of the true wavefunction. Thus

$$\langle \delta | \delta \rangle \leq \sum_{\alpha} |c_{\alpha}|^2 |e^{-iE_{\alpha}(t-t_0)} - \sum_{n=0}^{N_{\text{max}}} \frac{(-iE_{\alpha})^n}{n!} (t - t_0)^n|^2 \quad (21)$$

and the condition for the error to remain small is simply that $E_{\text{cutoff}} \Delta t < 1$. For a problem with $E_{\text{cutoff}} \ll E_H$, the least action propagator offers the possibility of much larger stepsizes at high accuracy than the Lanczos propagator.

The two propagators were tested numerically using a 1 dimensional Coulomb potential $1/x$ for x ranging from 0 to 10. The region was separated into 100 finite element regions, with two quadratic finite elements per region. The wavefunction was restricted to have zero value at both endpoints. The largest eigenvalue of the resulting Hamiltonian matrix was 499 Hartree. The initial wavefunction was chosen to be a Gaussian of unit width, centered at $x=2$. The choice of initial wavefunction and potential were made to ensure that the wavefunction would be far from equilibrium and have a strong interaction with the Coulomb potential, as for an electron wavepacket scattering from a positive ion.

The accuracy of the Lanczos- and least action propagator was calculated by propagating the initial wavefunction a single timestep and comparing the resulting wavefunction with the “true” wavefunction found by directly diagonalizing the full Hamiltonian. yielding an error

$$|err| = \langle \delta(t + \delta t) | \delta(t + \delta t) \rangle. \quad (22)$$

For the least action propagator, the order of propagation is one less than the degree of the polynomial basis functions. For the Lanczos propagator, the order is given by

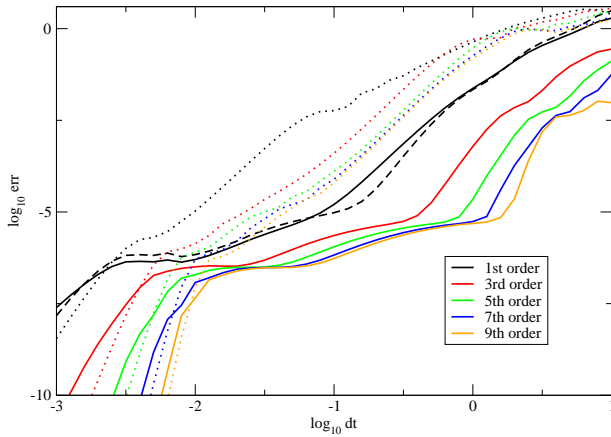


FIG. 1. (Color online) Point error $\langle \delta(x, t + \Delta t) | \delta(x, t + \Delta t) \rangle$ of Crank Nicholson (dashed line), Lanczos (dotted lines) and least action (solid lines) propagators, as a function of step size. For large stepsizes, the least action propagator is many times more accurate than the Lanczos propagator of the same order.

the dimension of the Krylov subspace, starting with 0 for the initial wavefunction.

The error as a function of order and stepsize for the two methods is shown in Figure 1. Also shown in the figure is the error vs time for the popular Crank Nicholson first order propagator [4]. As opposed to the global error which was used in the derivation of the least action propagator, these figures show the point error after a single propagation step.

These results show that the least action propagator offers the potential for large timesteps to be taken with high accuracy, with the greatest advantage coming from propagation at high order. At first order, the least action propagator gives approximately the same point error as the Crank Nicholson method, while higher orders rapidly decrease the error for a particular timestep, or alternatively increase the size of the timestep which can be taken for a particular desired error. As the order increases, the error begins to saturate as different order propagators converge on the same result. That this saturation does not result in zero error may result from numerical error in the diagonalization of the Hamiltonian or the linear solver.

For all orders tested, the least action propagator was many times more accurate than the Lanczos propagator for large stepsizes. For very small stepsizes, the error of both methods was comparable, with the Lanczos method more accurate. Both methods became much more accurate for stepsizes of less than 10^{-2} , which is interpreted to mean that the initial wavefunction had some projection onto very high energy eigenstates; ie, the sample problem did not have $E_{\text{cutoff}} \ll E_H$.

One weakness of the least action propagators which

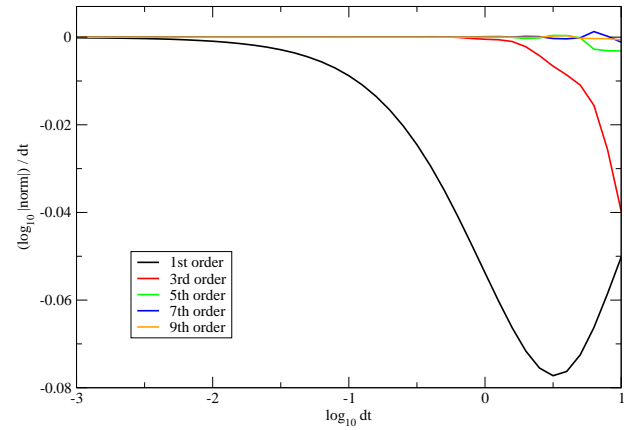


FIG. 2. (Color online) Exponential growth rate $(\log_{10} |\text{norm}|) / dt$ vs propagation time for different orders of the least action propagator. Higher orders show a growth rate closer to zero.

arises from the choice of polynomial basis functions is that the norm of the propagated wavefunction is not required to be a constant as a function of time. Figure 2 shows the rate of growth/decay of the norm $\langle \phi | \phi \rangle$ for different orders of propagation as a function of step size. Here the largest deviation from zero growth rate is found for the combination of low order and large stepsize. Higher order propagators show growth rates close to zero. For problems requiring repeated use of the propagator over many timesteps, it is thus likely that the propagated wavefunction will need to be renormalized periodically. Because the norm is still very close to 1, such renormalization has a minimal effect on the point error shown in Figure 1.

From these figures, it is apparent that the least action propagator works best at high orders, which offer the combination of large time steps, high accuracy and low rates of growth or decay of the norm. However, high order also increases the size of the linear system which must be solved at each step. For a basis consisting of n_x spatial and n_t temporal basis functions, the least action propagator requires $n_s = n_x(n_t + 1)$ variables to be solved for. While specific implementations are beyond the scope of this paper, the question of how best to solve this linear system will play a crucial role in applying the least action propagator to real world problems. To this end, a few features of the least action linear system are worth noting. First among these is the highly separable nature of the least action linear system defined in equations 14 and 15. In equation 14, the variation of the action is given as the sum of three matrices, $O_{ij}Q_{nm}$, $H_{ij}U_{nm}$, and V_{ijnm} . Of these, the first two are separated into the product of spatial and temporal matrices. Because of this, these two matrices inherit the sparsity and/or banded structure of the underlying spatial matrices. The case is similar for

the nonseparable V_{ijnm} : the integral over time and space is nonzero only if the integral over space is nonzero. Because of this, basis sets such as finite elements which are chosen for the structure of their Hamiltonian and overlap matrices will retain these advantages in the least action equation. For a banded problem such as the 1D finite element problem treated in this paper, the bandwidth of the linear system increases linearly with the order of the propagator, giving an overall n_t^2 scaling with the order.

For very large problems which lack such a simple structure, it is likely that solution of the least action linear system will require use of an iterative solver, such as those available in the PETSc [16] or Trilinos [17] libraries. Such solvers, require calculating a matrix vector product at every iteration. Here the advantages of the least action equation's separable form are very apparent, particularly in the case of a static Hamiltonian. A single matrix vector product of the linear system defined in equations 14 and 15 requires 1 (very expensive) matrix-vector multiplication by the unseparated matrix V_{ijnm} , $2n_t$ independent (expensive) multiplications of the form $D_{im} = M_{ij}C_{jm}$, where $M = O$ or H , followed by $2n_x$ (cheap) independent multiplications of the form $F_{in} = N_{nm}D_{im}$, where $N = Q$ or U . Thus, although the linear system which must be solved is very large, it is well suited to iterative solution. The overall scaling of one iteration with respect to propagator order will be limited by the slowest of these steps, which may depend on specifics of the data structure and architecture of the system used.

This paper has addressed the problem of propagating a wavefunction in time by minimizing the accumulated action. For a particular choice of spatial and temporal basis functions, the problem is reduced to solution of a (potentially very large) system of linear equations. This linear system inherits the sparsity and/or banded structure of the spatial Hamiltonian and overlap matrices, while its separable structure makes it amenable to solution by iterative solvers. The resulting propagator was shown to have improved convergence relative to the commonly used short iterative Lanczos propagator, giv-

ing the potential for larger stepsizes at high accuracy.

The derivation of the action as a measure of propagation error is a powerful result which offers many opportunities for the systematic improvement of propagation schemes. By monitoring the spacetime volumes where the most action is accumulated, spatial or temporal bases could be selectively refined to increase the total accuracy of a propagation step at minimal additional computational effort. For this reason, the full power of the action minimization principle may be yet to be unlocked.

-
- [1] Taylor, K. et al., *Journal of Electron Spectroscopy and Related Phenomena* **144** (2005) 1191.
 - [2] Feist, J. et al., *Physical review letters* **103** (2009) 63002.
 - [3] Sansone, G. et al., *Nature* **465** 763.
 - [4] Crank, J. and Nicolson, P., *Advances in Computational Mathematics* **6** (1996) 207.
 - [5] Askar, A. and Cakmak, A., *The Journal of Chemical Physics* **68** (1978) 2794.
 - [6] Bandrauk, A. and Shen, H., *Chemical Physics Letters* **176** (1991) 428.
 - [7] Park, T. and Light, J., *The Journal of chemical physics* **85** (1986) 5870.
 - [8] Tal-Ezer, H. and Kosloff, R., *The Journal of Chemical Physics* **81** (1984) 3967.
 - [9] Moler, C. and Van Loan, C., *SIAM review* **45** (2003) 3.
 - [10] Leforestier, C. et al., *Journal of computational physics* **94** (1991) 59.
 - [11] Truong, T. et al., *The Journal of Chemical Physics* **96** (1992) 2077.
 - [12] Kosloff, R., *The Journal of Physical Chemistry* **92** (1988) 2087.
 - [13] Triozon, F., Roche, S., and Mayou, D., *Riken Review* (2000) 73.
 - [14] Tal-Ezer, H., Kosloff, R., and Cerjan, C., *Journal of Computational Physics* **100** (1992) 179.
 - [15] Chen, R. and Guo, H., *The Journal of Chemical Physics* **119** (2003) 5762.
 - [16] Balay, S. et al., (2003).
 - [17] Heroux, M. et al., *ACM Transactions on Mathematical Software (TOMS)* **31** (2005) 397.